Current Topics in Genetics
Vol. 5, 2012

Review

Regulatory genomics: Insights from the zebrafish

Sumantra Chatterjee[$] and Thomas Lufkin[*,#]
Stem Cell and Developmental Biology, Genome Institute of Singapore, 60 Biopolis Street, 138672, Singapore

ABSTRACT

The sequencing of many vertebrate species over the last decade has opened up the possibility of using comparative genomics as a powerful tool to elucidate regulatory elements in the vertebrate genome. The zebrafish has played a pivotal role in this process. Its genome has been used in large-scale genome comparisons to locate vertebrate specific regulatory elements and also it has been an excellent model system to test out the predicted DNA sequences for their ability to drive reporter gene expression *in vivo*. In spite of all the successes there have still been some issues in using the zebrafish as a model system for these kinds of assays. This review will shed some light on the successes and failures of the zebrafish in pushing forward regulatory genomics.

KEYWORDS: zebrafish, *cis*-regulatory elements, bacterial artificial chromosomes (BACs), histone modifications

INTRODUCTION

The complete sequencing of the human genome [1, 2] and genomes of other species [3, 4] has opened up an unprecedented possibility that the genomes can be explored in greater detail to uncover all its facets. Though there has been substantial progress in deciphering the total number of protein coding genes in the human genome [5] the search of all the regulatory elements in various vertebrate genomes has been a challenge. The ENCODE project [6] is a large scale effort in annotating such functional elements in the human genome and have been able to decipher many such elements but is by no means an exhaustive survey. Functional non-coding elements in the genome include non-coding RNA, *cis*-regulatory elements, splicing elements and sequences directing chromatin structure. Unlike protein coding genes which have a characteristic structure that helps in their identification, very little is known about the organization of the non-coding elements and hence makes it challenging to locate and validate them. Multiple approaches have been employed to locate *cis*-regulatory elements. The traditional or pre-genomic era techniques to locate *cis*-regulatory elements primarily relied on biochemical and genetic assays to capture them. One of the earliest biochemical methods to study *cis*-regulatory elements was DNA footprinting [7]. Another biochemical assay, which captures the state of chromatin *in vivo* is the DNase I hypersensitivity assay [8, 9].

The availability of the various genome sequences and the concurrent development of genomic alignment, visualization, and analytical bioinformatics tools have made comparative genomics not only possible but also an increasingly popular approach for the discovery of putative *cis*-regulatory elements. Comparative genomics is not biased towards any genomic region and does not rely on any prior knowledge of which transcription factors are regulating the target gene. The basic

*Corresponding author
[#]lufkin@gis.a-star.edu.sg
[$]chatterjees@gis.a-star.edu.sg

premise of this method is that functional non-coding DNA tends to evolve more slowly than non-functional DNA due to selective pressure, hence *cis*-regulatory elements can be identified as conserved non-coding elements in sequence comparison of related genomes. Many such conserved non-coding elements (CNEs) lie next to critical developmental control genes and have been shown to be developmental enhancers when selected and tested in mice and zebrafish transgenic assays [10-12]. In the post genome sequence era, chromatin immunoprecipitation followed by massive sequencing (ChIP-Seq) for P300 and ChIP followed by microarray (ChIP-Chip) on histone modifying mark H3K4me3 (Histone 3 Lysine 4 trimethylation) and histone acetylation have been shown to be useful in identifying tissue specific enhancers on a global scale, many of which have very weak sequence conservation [13-18].

Most of the methods described above help to locate putative regulatory elements in the genome, but to be exactly certain of its role, functional validation is a must. Functional validation helps to segregate binding sites (DNA fragments) which direct spatio-temporal expression of neighboring genes from binding sites which might not have a direct role in gene regulation and this is critical in our understanding of transcriptional control and goes a long way in helping to build gene regulatory networks. Though some studies have utilized the more rapid luciferase assay in cell lines to validate the functionality of a binding site [15, 19, 20], *in vivo* transgenics is still a more powerful and convincing method for such validations, especially if working with developmental control genes. This review will look at various methods employed to utilize the zebrafish in both locating and validating *cis*-regulatory elements.

Fish comparative genomics

The two main methods employed to locate conserved functional enhancers in the genome are either through whole genome comparisons or by locus-by-locus alignment. Whole genome alignments typically use local alignment programs like BLASTZ [21] and MegaBLAST [22] to rapidly align regions of high homology. When such an alignment is carried out between phylogenetically distant species such as human and fish they generally fail to identify and align all the orthologous sequences due to stringent criteria that are set for local alignments. On the other hand, locus by locus global alignments of orthologous gene loci using programs like LAGAN/MLAGAN [23] and AVID [24] are effective in identifying all the associated CNEs. But it is prudent to bear in mind that since global alignment has an additional assumption that input sequences occur in the same order and orientation, they can detect more weakly conserved regions [24, 25]. Nevertheless, it should be noted that global alignment algorithms tend to miss out on conserved functional elements that have undergone local inversions or rearrangements.

While human-mouse sequence conservation can be particularly useful for finding mammalian specific *cis*-regulatory elements, this approach tends to identify many false positives due to the relatively short evolutionary distance between the two species (~60 Myr). On the other hand, as the phylogenetic distance between two species increases the average conservation of the DNA under neutral evolution since their last common ancestor decreases. Sequence comparison between human and more distantly related species like teleost fishes that have been separated by larger evolutionary distance (~450 Myr), increases the probability that the CNEs that are identified are functional ones [26]. The first proof of principle study to show the feasibility of comparative genomics between human and teleost fish was reported in 1995. Two transcriptional enhancers that control *Hoxb4* expression in the mouse mesoderm, ectoderm and the neural tube were found to contain three mouse-fugu conserved regions (CR1, CR2 and CR3). CR1 was shown to be essential for expression in the mesoderm, central and peripheral nervous system while CR3 directed gene expression to the posterior hindbrain [27]. Mouse-fugu comparison has also located *cis*-regulatory elements for the *Pax9/Nkx2-9* locus. Out of a total of 15 CNEs in the mouse-human comparison 2 CNEs were also present in the human-mouse-fugu comparison, one of which directed gene expression in ventral neural tube for *Nkx2-9* and the other in the medial nasal processes for *Pax9* [28]. Human-fugu

comparison also revealed that seven out of nine CNEs were functional in the gene desert surrounding the *Dachshund* gene locus [29]. Human-fish comparison has also been done to locate CNEs in the whole genome rather than on a gene-by-gene basis and have shown that many of these sequences can recapitulate gene expression *in vivo*, thus further strengthening the argument that phylogenetically distant conserved non coding elements have a higher chance of having a function [11, 30]. But a recent study looking at human-mouse-zebrafish sequence constraints has also reported the presence of many conserved elements which are non functional as enhancers [31], highlighting some of the issues in only using sequence conservation as a yardstick for enhancer function.

In vivo validation by small vectors

The process of functional validation of various CNEs by the small vector method is mainly "targeted", where the PCR amplified CNE DNA is either ligated to a reporter vector [11] or in the case of zebrafish is co-injected with the reporter construct [12, 32, 33], and analyzed for transient expression of green fluorescent protein (GFP) or beta-galactosidase encoded by the *lacZ* gene. This expression of the CNE driven reporter molecule is then matched back to the expression domains of the nearby genes around which the CNE was located to see overlapping expression leading to association of the CNE (enhancer) to a gene. There are a couple of drawbacks to this system as well, since the integration of the transgene occurs at random location in the genome, there is always a risk of the element showing ectopic expression due to it coming under the control of an endogenous *cis*-regulatory element or it being silenced (positional effects). Hence it is crucial that similar expression is obtained in several independent transgenic lines before drawing a conclusion on CNE activity. Secondly, the microinjection of the vector is carried out at the one cell stage so that the transgene is integrated into the genome before the cell divides and hence all cells in the resulting embryo will contain the transgene. But in certain cases especially in the rapidly dividing zebrafish embryo there is always the chance of the integration happening after cell division has started, resulting in a mosaic expression of the transgene. But the biggest advantage in zebrafish is the availability of a large number of embryos, which can be injected with the same construct and hence screened to detect the actual expression domain over the mosaic expression. Since the embryos are translucent, using fluorescent reporters like EGFP one can monitor the transgene expression pattern over the course of development in each animal. Also individual transgenic embryos can be sectioned and stained for specific antibody (EGFP, beta-galactosidase) to detect the exact tissue in which the enhancer expresses [31, 34].

Another approach uses the Tol2 transposon system that allows CNE-reporter gene fusions to be integrated into the germline more efficiently thus achieving much more robust expression of the transgene and solves the problem of mosaicism associated with transient transgenics in zebrafish [35-37]. Over the years there has been marked improvement in various Tol2 based vectors to make this even more robust. The Zebrafish Enhancer Detector (ZED) vector harbors several key improvements, among them a sensitive and specific minimal promoter chosen for optimal enhancer activity detection, insulator sequences to shield the minimal promoter from position effects, and a positive control for transgenesis [38]. The Tol2 mediated transgenesis is the favored method for testing such regulatory elements now due to its relative lack of mosaicism.

The biggest advantage that is conferred by using the fish for transgenic assays is the high-throughput. The fish undergoes external fertilization and a single mating pair produces numerous embryos. This allows for numerous DNA elements to be tested at once and their activity monitored in a dish over the course of development of the fish. The development cycle of the fish is also faster and in 72 hours post fertilization (hpf) most organs and organ systems are formed in the fish. In contrast the mouse is a much more expensive model to use for such functional validation both due to the limited number of embryos available as well as its in-utero development.

Zebrafish in the world of conservation

It has now been clearly demonstrated that high levels of functional conservation of genes is not

necessarily associated with sequence conservation of their corresponding regulatory elements. Moreover, highly conserved non-coding elements may not have a known role in *cis*-regulation and relying on standard methods to detect evolutionary constraint overlooks significant functional information. But more and more evidence is gathering that the loss of function of these conserved elements is a genome wide phenomenon between fishes and other vertebrates. A study reported that only 10% of 104 mouse enhancers experimentally validated have homologous sequences in zebrafish [39], and several reports have suggested that the genome has changed rapidly in the teleost lineage. There was whole genome duplication (WGD) very early in the teleost radiation [40, 41]. The duplication would have brought about radical remodeling in the teleost genome, being accompanied with gene degeneration and complementation [42]. Thus the loss of activity or the presence of enhancer elements that are not constrained in sequence between zebrafish and mammals is not surprising. Studies by groups looking at single genes at a time have also started revealing similar trends in other genomic loci in the zebrafish [43, 44]. But there is a need to interpret these data with caution as the teleost radiation occurred 300-400 million years ago and thus, zebrafish and fugu are separated by a much larger evolutionary distance than mammals and this may also, in part, explain the lack of observable alignment or constraint.

Bacterial artificial chromosomes (BACs) to detect gene regulation

In vivo transgenic assays have clearly demonstrated the need for "context" for regulatory elements to function in a spatio-temporal manner. In one study the authors located an enhancer for the amyloid precursor protein gene (*APPb*) within the intron of the gene [45]. Although the enhancer was active in specific non-neural cells of the notochord when placed with the endogenous gene's promoter, its function was restricted to specific expression in neurons when juxtaposed with additional far upstream promoter elements of the gene. The authors demonstrated that expression of GFP fluorescence resembling the tissue distribution of *APPb* mRNA requires both the intron enhancer and ~28 kb of DNA upstream of the gene. The results indicate that tissue-specificity of an enhancer tested out of its native context may be quite different from that in the context of its own gene. Thus it is crucial that enhancer elements should be tested for function in their native context with surrounding DNA sequences to capture their complete profile and hence the requirement for special vectors to clone and capture large stretches of genomic DNA for *in vivo* testing and creating transgenics.

Bacterial artificial chromosomes (BACs) are such specialized vectors that can be used to clone in large DNA inserts that range in size from 150-300 kb. Large chunks of contiguous genomic sequences have been cloned into BAC vectors and have been used for sequencing of various genomes. Hence BACs are useful tools to modify a gene and its regulatory elements in their genomic context by introducing a reporter construct into the vector backbone of BACs. When assayed in transgenic animals, any enhancers present in the genomic insert of the BAC should be able to stimulate the gene promoter and drive transcription of the reporter gene, yielding an *in vivo* readout of any enhancer activity harbored within a particular BAC. Hence, by comparing the expression of the reporter to the expression of the endogenous gene allows for the determination of the regulatory elements for the gene that are present in the genomic region cloned in the BAC [46, 47]. Most genes in the vertebrate genomes have regulatory elements that are scattered over megabases around the gene and hence to capture all the *cis*-regulatory elements will require extensive cloning, which is time consuming and in the end may still fail to detect all the functional enhancers. In a recent study the authors demonstrated that for developmental genes in zebrafish modified BACs can be used to uncover regulatory elements which not only are distal from the TSS of the gene but were also not captured by just looking at conserved sequences [31]. This study showed the utility of using large regions of the genome to detect and validate enhancers for genes whose regulatory domains are spread over large stretches of the genome. The modified BAC can be further fragmented into smaller pieces of DNA to zoom into the actual regulatory sequence. The modified BAC method is not based on the

need for any specific histone marks or genomic features like sequence conservation and thus allows for an unbiased assay of the genome to locate *cis*-regulatory elements. For a review on BAC modifications in zebrafish see [48].

Various methods for BAC modification

Homologous recombination mediated BAC transgenics

One of the earliest methods of BAC modifications was developed by Copeland [49] and Stewart [46, 47]. The method relies on the ability to insert a reporter gene next to the promoter of a gene by homologous recombination in bacteria. This allows for the generation of BACs which when injected into 1-cell stage zebrafish embryos allows for the visualization of all domains of expression of the gene as controlled by its regulatory elements present in the BAC. Like any method BAC modification by homologous recombination has some limitations. In zebrafish the rate of germ line transmission of these BACs is very low, compared to the smaller vectors. This increases the time needed to raise stable lines for further studies. Another limitation of these large BAC transgenes, which they share with small vectors, is their susceptibility to genomic silencing and perturbations in genomic landscapes that could confer spurious promoter/cis-regulatory module (CRM) activities. These BACs also form concatamers inside the embryo and concatemeric transgenes many times lead to silencing, instability, and genetic lesions [50, 51].

Transposon mediated BAC transgenics

Tol2 transposon based BAC modification has recently been employed in zebrafish [52]. One of the advantages of using this method is that only a single copy of the modified BAC is delivered per transgenic embryo. This method has an added step to introduce a *Tol2* cassette containing the minimal *cis*-sequences of *Tol2* in an inverted orientation separated by a ~1 kb spacer. This cassette enables incorporation of the *Tol2 cis*-sequences essential for transposition into a BAC clone through a single step of homologous recombination. Along with single copy insertion of the transgene the other obvious advantage is the absence of rearrangements of the transgene within the embryo thus not compounding the observed expression domains.

Histone modifications and regulatory genomics

Over the last decade it has been demonstrated that a number of post-translationally modified forms of histone protein subunits associate with distinct *cis*-regulatory elements. For example, the trimethylated lysine 4 of histone 3 (H3K4me3) is associated with genes that are actively transcribed [53, 54]. Large-scale and genome-wide profiling studies of H3K4me3 in the human and mouse genomes indicate that this modification is preferentially associated with the promoters of active genes [55, 56] and, to a lesser degree, at distal *cis*-regulatory elements, such as enhancers [55]. Mono methylated lysine 4 of histone 3(H3K4me1) is similarly associated with active enhancer elements as well as transcriptional start sites, though to a lesser degree than H3K4me3 [15, 55]. H3K4me1 is also found at insulator elements in association with binding by CTCF [55, 57]. While H3K4me3, and to some degree H3K4me1, are generally activation marks, other histone marks are indicative of gene repression. ChIP-on-chip has been applied to early zebrafish embryos to investigate the binding of modified histones within the genome. Initial studies proved that H3K4me3 binding was enriched at transcriptional start sites similar to findings in mammalian cells [58]. A similar approach was used to demonstrate that both H3K4me3 and H3K27me3 marks are deposited at developmentally regulated genes independent of transcriptional activation immediately prior to the onset of zygotic transcription in zebrafish [59].

By applying chromatin immunoprecipitation followed by deep sequencing, recently Aday *et al*. [60] have demonstrated that H3K4me1 and H3K4me3 are enriched at transcriptional start sites in the genome of the developing zebrafish embryo and that this association strongly correlates with gene expression. They further demonstrate that these modifications associate with distal non-coding conserved elements, including known active enhancers, thus showing the utility of using these histone marks at different stages of zebrafish development to elucidate most of the key transcriptional

enhancers. A very recent study has shown that in zebrafish there is an epigenetic prepatterning of developmental gene expression. This involves pre-zygotic gene activation (ZGA) marking of transcriptionally inactive genes involved in homeostatic and developmental regulation by permissive H3K4me3 with or without repressive H3K9me3 or H3K27me3 [61]. In zebrafish, pre-ZGA development for ten cell cycles provides an opportunity to examine whether genomic enrichment in modified histones is present before initiation of transcription thus allowing for detailed elucidation of this phenomenon which would have been difficult in mammals where the ZGA occurs very early in development. It is becoming abundantly clear now that with the advances in sequencing technologies and the lowering cost of sequencing the elucidation of regulatory elements in the future will solely be based on unbiased high throughput sequencing, looking both at histone modification marks as well as specific transcription factor binding sites.

CONCLUDING REMARKS

The zebrafish has been an outstanding ally for geneticist over the last 2 decades in deciphering gene regulation and regulatory networks. Its many advantages as a model organism have been used by researchers to circumvent some of the shortcomings of higher vertebrates. Though like any model organism it has its own disadvantages, including a partially duplicated vertebrate genome and lack of certain mammalian organs (e.g. uterus, mammary glands, hair), yet its usefulness in genomics has been highlighted by the ability to do both reverse and forward genetics screen as well as ChIP for its transcription factors and histone modifying marks. The rapid division cycle as well as availability of numerous embryos have ensured that the fish can be used in greater experimental numbers than that would be possible with the mouse. Numerous studies have also demonstrated that the fish can be a highly useful organism to screen large number of potential regulatory elements, thus acting as a screening tool. With increasing availability of good antibodies and decreasing cost of sequencing, more and more transcription factor binding sites (potential regulatory elements) will be discovered rapidly in the fish and owing to its close similarity with the human genome, would potentially be a good starting point in any human specific study in the future.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T.,

Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. J. 2001, Nature, 409, 6822.

2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K.,

Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. 2001, Science, 291, 5507.

3. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., and Venter, J. C. 2000, Science, 287, 5461.

4. Woods, I. G., Kelly, P. D., Chu, F., Ngo-Hazelett, P., Yan, Y. L., Huang, H., Postlethwait, J. H., and Talbot, W. S. 2000, Genome Res., 10, 12.

5. Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., and Lander, E. S. 2007, Proc. Natl. Acad. Sci. USA, 104, 49.

6. Myers, R. M., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R. C., Bernstein, B. E., Gingeras, T. R., Kent, W. J., Birney, E., Wold, B., and Crawford, G. E. 2011, PLoS Biol., 9, 4.

7. Galas, D. J. and Schmitz, A. 1978, Nucleic Acids Res., 5, 9.

8. Keene, M. A., Corces, V., Lowenhaupt, K., and Elgin, S. C. 1981, Proc. Natl. Acad. Sci. USA, 78, 1.

9. McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D., and Felsenfeld, G. 1981, Cell, 27, 1 Pt 2.
10. Sandelin, A., Bailey, P., Bruce, S., Engstrom, P. G., Klos, J. M., Wasserman, W. W., Ericson, J., and Lenhard, B. 2004, BMC Genomics, 5, 1.
11. Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., and Rubin, E. M. 2006, Nature, 444, 7118.
12. Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J., Cooke, J. E., and Elgar, G. 2005, PLoS Biol., 3, 1.
13. Blow, M. J., McCulley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Bristow, J., Ren, B., Black, B. L., Rubin, E. M., Visel, A., and Pennacchio, L. A. 2010, Nat. Genet., 42, 9.
14. Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., and Pennacchio, L. A. 2009, Nature, 457, 7231.
15. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. 2007, Nat Genet., 39, 3.
16. Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., and Ren, B. 2009, Nature, 459, 7243.
17. Roh, T. Y., Wei, G., Farrell, C. M., and Zhao, K. 2007, Genome Res., 17, 1.
18. Roh, T. Y., Cuddapah, S., and Zhao, K. 2005, Genes Dev., 19, 5.
19. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C.,Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W. K., Clarke, N. D., Wei, C. L., and Ng, H. H. 2008, Cell, 133, 6.
20. Petrykowska, H. M., Vockley, C. M., and Elnitski, L. 2008, Genome Res., 18, 8.
21. Schwartz, S., Kent, W. J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R. C., Haussler, D., and Miller, W. 2003, Genome Res., 13, 1.
22. Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000, J. Comput. Biol., 7, 1-2.
23. Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., Green, E. D., Sidow, A., and Batzoglou, S. 2003, Genome Res., 13, 4.
24. Bray, N., Dubchak, I., and Pachter, L. 2003, Genome Res., 13, 1.
25. Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I., and Hardison, R. C. 2003, Genome Res., 13, 1.
26. Elnitski, L., Hardison, R. C., Li, J., Yang, S., Kolbe, D., Eswara, P., O'Connor, M. J., Schwartz, S., Miller, W., and Chiaromonte, F. 2003, Genome Res., 13, 1.
27. Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R., and Brenner, S. 1995, Proc. Natl. Acad. Sci. USA, 92, 5.
28. Santagati, F., Abe, K., Schmidt, V., Schmitt-John, T., Suzuki, M., Yamamura, K., and Imai, K. 2003, Genetics, 165, 1.
29. Nobrega, M. A., Ovcharenko, I., Afzal, V., and Rubin, E. M. 2003, Science, 302, 5644.
30. Shin, J. T., Priest, J. R., Ovcharenko, I., Ronco, A., Moore, R. K., Burns, C. G., and MacRae, C. A. 2005, Nucleic Acids Res., 33, 17.
31. Chatterjee, S., Bourque, G., and Lufkin, T. 2011, BMC Dev. Biol., 11, 63.
32. Muller, F., Williams, D. W., Kobolak, J., Gauvry, L., Goldspink, G., Orban, L., and Maclean, N. 1997, Mol. Reprod. Dev., 47, 4.
33. Goode, D. K., Callaway, H. A., Cerda, G. A., Lewis, K. E., and Elgar, G. 2011, Development, 138, 5.

34. Kikuta, H., Fredman, D., Rinkwitz, S., Lenhard, B., and Becker, T. S. 2007, Genome Biol., 8 Suppl 1.
35. Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., Urasaki, A., Kawakami, K., and McCallion, A. S. 2006, Nat. Protoc., 1, 3.
36. Suster, M. L., Kikuta, H., Urasaki, A., Asakawa, K., and Kawakami, K. 2009, Methods Mol. Biol., 561, 41.
37. Parinov, S., Kondrichin, I., Korzh, V., and Emelyanov, A. 2004, Dev. Dyn., 231, 2.
38. Bessa, J., Tena, J. J., de la Calle-Mustienes, E., Fernandez-Minan, A., Naranjo, S., Fernandez, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., and Gomez-Skarmeta, J. L. 2009, Dev. Dyn., 238, 9.
39. Plessy, C., Dickmeis, T., Chalmel, F., and Strahle, U. 2005, Trends Genet., 21, 4.
40. Postlethwait, J. H., Yan, Y. L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E., Paw, B., Ransom, D., Singer, A., Thomson, M., Abduljabbar, T. S., Yelick, P., Beier, D., Joly, J. S., Larhammar, D., Rosa, F., Westerfield, M., Zon, L. I., Johnson, S. L., and Talbot, W. S. 1998, Nat. Genet., 18, 4.
41. Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. 1998, Science, 282, 5394.
42. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. 1999, Genetics, 151, 4.
43. McGaughey, D. M., Vinton, R. M., Huynh, J., Al-Saif, A., Beer, M. A., and McCallion, A. S. 2008, Genome Res., 18, 2.
44. Antonellis, A., Huynh, J. L., Lee-Lin, S. Q., Vinton, R. M., Renaud, G., Loftus, S. K., Elliot, G., Wolfsberg, T. G., Green, E. D., McCallion, A. S., and Pavan, W. J. 2008, PLoS Genet., 4, 9.
45. Shakes, L. A., Malcolm, T. L., Allen, K. L., De, S., Harewood, K. R., and Chatterjee, P. K. 2008, Nucleic Acids Res., 36, 19.
46. Muyrers, J. P., Zhang, Y., Testa, G., and Stewart, A. F. 1999, Nucleic Acids Res., 27, 6.
47. Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. 1998, Nat. Genet., 20, 2.
48. Chatterjee, S. and Lufkin, T. 2011, Mol. Biosyst., 7, 8.
49. Lee, E. C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N. A., and Copeland, N. G. 2001, Genomics, 73, 1.
50. Garrick, D., Fiering, S., Martin, D. I., and Whitelaw, E. 1998, Nat. Genet., 18, 1.
51. Dorer, D. R. and Henikoff, S. 1997, Genetics, 147, 3.
52. Suster, M. L., Sumiyama, K., and Kawakami, K. 2009, BMC Genomics, 10, 477.
53. Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. 2002, Proc. Natl. Acad. Sci. USA, 99, 13.
54. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. 2002, Nature, 419, 6905.
55. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. 2007, Cell, 129, 4.
56. Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J. 3rd, Gingeras, T. R., Schreiber, S. L., and Lander, E. S. 2005, Cell, 120, 2.
57. Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., Francoijs, K. J., Stunnenberg, H. G., and Veenstra, G. J. 2009, Dev. Cell, 17, 3.
58. Wardle, F. C., Odom, D. T., Bell, G. W., Yuan, B., Danford, T. W., Wiellette, E. L., Herbolsheimer, E., Sive, H. L., Young, R. A., and Smith, J. C. 2006, Genome Biol., 7, 8.
59. Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., Rinn, J., and Schier, A. F. 2010, Nature, 464, 7290.
60. Aday, A. W., Zhu, L. J., Lakshmanan, A., Wang, J., and Lawson, N. D. 2011, Dev. Biol., 357, 2.
61. Lindeman, L. C., Andersen, I. S., Reiner, A. H., Li, N., Aanes, H., Ostrup, O., Winata, C., Mathavan, S., Muller, F., Alestrom, P., and Collas, P. 2011, Dev. Cell, 21, 6.